Urban Big Data Centre

Data and expertise for improving cities

www.ubdc.ac.uk

GIS-based facility location analysis for the public and private sectors

Transcript from webinar video recording

1

00:00:01,120 --> 00:00:07,280

[Muir Houston] Morning everyone and welcome to this session on GIS-based facility location analysis for the public

2

00:00:07,280 --> 00:00:15,120

and private sectors. In this webinar, led by UBDC's Dr Jing Yao and supported by Weicong Luo, we will

3

00:00:15,120 --> 00:00:21,440

introduce some basic classic location models with examples of location allocation analysis

4

00:00:21,440 --> 00:00:28,320

and ArcGIS using open data. Siting key facilities such as hospitals, schools, and fire stations in the

5

00:00:28,320 --> 00:00:33,920

right locations is vital for providing the services we need. Cities are continually reviewing the location

6

00:00:33,920 --> 00:00:39,840

of these facilities as populations grow or services are reorganised. New technologies such as

7

00:00:39,840 --> 00:00:45,760

electric or hydrogen power for vehicles can create demand for new networks of service centres within

8

00:00:45,760 --> 00:00:51,920

existing urban areas. Jing has a PhD in Geography and an MSc in Industrial Engineering both from

9

00:00:52,560 --> 00:01:00,880

Arizona State University. She also has an MSc and a BSc in Geographical Information Systems from Nanjing

10

00:01:00,880 --> 00:01:06,320

University. Jing's research interests cover a wide range of areas in geographic information science,

11

00:01:06,320 --> 00:01:12,880

including spatial analysis, spatial statistics, spatial modeling and spatial optimisation.

12

00:01:12,880 --> 00:01:17,840

Currently she focuses on developing quantitative methods for spatially integrated analysis and

13

00:01:17,840 --> 00:01:22,400

modeling and particularly their applications in social science and health research.

14

00:01:22,960 --> 00:01:28,160

I see with participants from Australia, Chile, China, Hungary, Nigeria, Turkey, the Philippines,

15

00:01:28,160 --> 00:01:34,240

Russia, Ukraine, USA, and the UK. Sorry if I've missed any of you. Now I'd just like to remind participants

16

00:01:34,240 --> 00:01:38,960

that the session is recorded and will be uploaded on the web in an accessible format at some point

17

00:01:38,960 --> 00:01:45,840

after the session. Details will be provided on the UBDC website. Also check out the website for

18

00:01:45,840 --> 00:01:52,080

other resources including how to access data and other training and events delivered by UBDC.

19

00:01:52,800 --> 00:01:57,920

Cameras will be turned off and microphones muted to aid privacy and also for bandwidth reasons.

20

00:01:57,920 --> 00:02:02,800

Feel free to introduce yourself in the chat box but please do not include personal information

21

00:02:02,800 --> 00:02:08,000

such as phone numbers or emails. Please use the question and answer facility to ask

22

00:02:08,000 --> 00:02:14,080

questions. These will be collated, and responses provided in the Q&A session. The presentation

23

00:02:14,080 --> 00:02:20,800

will last for 30 to 40 minutes and then we'll have a 10 to 15 minute long question and answer session.

24

00:02:21,520 --> 00:02:35,840

So, I'd just like to introduce Jing and hand over to her to let the session begin. Thank you.

25

00:02:44,560 --> 00:02:49,840

[Jing Yao] You see the screen now? [Weicong Luo] Yep. [Jing Yao] Ok, that's good.

26

00:02:56,000 --> 00:03:02,960

Morning everyone. So, thanks for attending today's webinar. I'm Jing Yao. I'm currently a

27

00:03:02,960 --> 00:03:08,800

Senior Lecturer at the Urban Big Data Centre at the University of Glasgow. So today

28

00:03:08,800 --> 00:03:15,040

I will talk about GIS-based facility location analysis for the public and private sectors.

29

00:03:17,680 --> 00:03:25,920

So, location analysis is essentially about where should a facility be and the task of location analysis

30

00:03:25,920 --> 00:03:34,080

is actually as old as our world. So, for example, in the early prehistoric times people needed

31

00:03:34,080 --> 00:03:39,040

to determine the best location for hunting food, harvesting natural resources and so on.

32

00:03:40,160 --> 00:03:47,920

So nowadays location is considered one of the most important factors that affect the success of a

33

00:03:47,920 --> 00:03:54,880

private or public sector organisation. So, for example in the private sector

34

00:03:54,880 --> 00:04:02,960

a good location for a coffee shop or retail store can bring a large number of customers

35

00:04:02,960 --> 00:04:09,680

and therefore generate high profits. So, in the public sector good locations

36

00:04:10,240 --> 00:04:16,720

of schools and hospitals can ensure social equity in accessing such public services.

37

00:04:19,680 --> 00:04:23,920

So here are some more examples of facility location problems.

38

00:04:24,960 --> 00:04:32,080

First, what is the minimum number of fire stations and where should they be located?

39

00:04:34,240 --> 00:04:41,360

Where should they be located necessary to cover 90 percent of building fires within 4 minutes of a call for

40

00:04:41,360 --> 00:04:50,960

service? And for a manufacturing plant, where should it be located so that the total

41

00:04:50,960 --> 00:04:58,400

transportation costs can be minimised? So, if Glasgow City Council wishes to close

42

00:04:58,400 --> 00:05:04,240

11 of their existing 148 primary schools, which schools should be closed?

43

00:05:06,800 --> 00:05:14,240

So, solving facility location problems is not easy. So, think about locating a retail store.

44

00:05:14,240 --> 00:05:24,080

So, assume we want to pick a site which will maximise the economic return. And this return

45

00:05:24,080 --> 00:05:29,520

can be affected by many factors. For example, the number of potential customers within market

46

00:05:29,520 --> 00:05:41,760

area, accessibility of the site (is it on a main street? Is it possible to turn left into the site?) and also affected by the visibility of the store

47

00:05:44,320 --> 00:05:55,800

the appearance of store (is it attractive?) and how much it will cost to build the site - to arrange to buy the site and also the cost of construction.

48

00:05:58,000 --> 00:06:04,800

For now, let's think about the possible number of solutions we can have for facility location

49

00:06:04,800 --> 00:06:12,560

problems. So, think about to select *p* facilities from *n* facilities we can use this

50

00:06:12,560 --> 00:06:21,760

mathematical formula to calculate the number of all possible solutions. And if we again

51

00:06:22,480 --> 00:06:32,000

take a Glasgow primary schools as an example. If we want to select 11 out of existing 148 schools,

52

00:06:32,640 --> 00:06:41,760

how many solutions we can have? So, we can see this is a very large number so it

53

00:06:41,760 --> 00:06:51,280

is very difficult to solve the problem just by enumeration and therefore we need a more efficient

54

00:06:51,280 --> 00:06:57,120

way to solve facility location problems, which is the optimisation models I will talk about next.

55

00:07:01,680 --> 00:07:11,760

So, we usually formulate the facility location problems by optimisation models. So, an

56

00:07:11,760 --> 00:07:18,480

optimisation model has three components: objectives, constraints, and decision variables.

57

00:07:19,520 --> 00:07:24,080

So, an objective is the goal we want to achieve.

58

00:07:24,720 --> 00:07:30,240

For example, we want to maximise the market share, or we want to minimise the total transportation costs.

59

00:07:31,360 --> 00:07:37,360

Achieving such objectives usually are subject to budget or resource constraint.

60

00:07:38,800 --> 00:07:44,240

Decision variables are the decisions we are going to make. For example, whether we are going

61

00:07:44,240 --> 00:07:50,480

to site a facility at a particular location or whether we should close a particular school.

62

00:07:53,440 --> 00:07:56,960

So, we will further explain those three components

63

00:07:57,920 --> 00:08:02,400

through three classic facility location models.

64

00:08:04,560 --> 00:08:13,920

the first is *p*-median problem. It is to locate a multiple number of facilities and allocate

65

00:08:13,920 --> 00:08:19,600

the demand served by these facilities so that the system is as efficient as possible.

66

00:08:21,040 --> 00:08:28,720

So, we have an example here. So, in this example we assume facilities are warehouses and demands

67

00:08:28,720 --> 00:08:35,520

are outlets. So, each outlet target has a weight so we use 'W' to represent

68

00:08:35,520 --> 00:08:41,120

the weight. So, this weight can represent the amount of goods needed from the warehouse.

69

00:08:42,560 --> 00:08:48,880

And we can assume *p* is 4 so, the problem here would be

70

00:08:48,880 --> 00:08:56,080

where should we site four warehouses to serve all the outlets

71

00:08:56,080 --> 00:09:02,720

so that the total transportation costs are minimised.

72

00:09:03,600 --> 00:09:10,560

So here we define the system efficiency as the total weighted transportation costs

73

00:09:10,560 --> 00:09:16,640

and this cost can be measured by travel distance or travel time between the warehouses and the

74

00:09:16,640 --> 00:09:25,360

outlets. So, each outlet is served by the nearest warehouse. So, this is the median problem.

75

00:09:27,520 --> 00:09:33,680

This is the mathematical formulation of the median problem. First, we look at the notation. i and j

76

00:09:33,680 --> 00:09:42,640

are to represent the demand areas and facility locations and d_{ij} represents the

77

00:09:42,640 --> 00:09:48,880

shortest distance or travel time between demand area i and facility j. So, W_i represents

78

00:09:48,880 --> 00:09:54,480

the amount of demand in demand area i. So, p is the total number of facilities to be located.

79

00:09:55,600 --> 00:10:02,480

Here we have two types of decision variables. Y_j indicates whether we should site a facility

80

00:10:02,480 --> 00:10:08,960

or not. X_{ij} indicates whether the demand area i is served by facility j or not.

81

00:10:10,960 --> 00:10:16,320

So, for the model on the right-hand side, we can see the objective is to minimise the total

82

00:10:16,320 --> 00:10:22,960

weightage travel distance, travel transportation cost. So here we use the travel

83

00:10:22,960 --> 00:10:32,480

distance or travel time to represent the cost and we use this to represent the system efficiency. So

84

00:10:32,480 --> 00:10:40,480

actually, when we minimise the total travel distance or time actually we maximise

85

00:10:40,480 --> 00:10:47,280

the accessibility here. So, we can see when we minimise the weight here the travel distance time

86

00:10:47,280 --> 00:10:56,000

actually will maximise that accessibility. For constraints, the constraint one ensures each

87

00:10:56,000 --> 00:11:02,240

demand area is to be served by exactly one facility, actually the nearest facility. And

88

00:11:03,040 --> 00:11:10,000

constraint two ensures no allocation occurs unless a facility is sited. So, we can see if the value of

89

00:11:10,000 --> 00:11:19,360

X_{ij} is 0, which means demand *i* is not solved by facility *j*. So, in that case the value of Y_j can be 1 or

90

00:11:19,360 --> 00:11:26,560

0, which means it doesn't matter whether facility *j* is located or not. But if the value of X_{ij} is 1,

91

00:11:26,560 --> 00:11:33,280

which means that the demand area *i* is served by facility *j*. So, this means that the value of Y_j

92

00:11:33,280 --> 00:11:45,440

must be one that means facility *j* must be located. So, constraint three indicates that the total number of facilities

93

00:11:45,440 --> 00:11:55,840

to be located is *p*. Constraint 4 requires that the values of X_{ij} and Y_j only can be zero or one.

94

00:11:59,120 --> 00:12:07,520

So, this is an example of optimising healthcare facilities and resource allocation in Gaza

95

00:12:07,520 --> 00:12:17,680

Province, Mozambique. So, this is an empirical study I have done before. The purpose is to evaluate the locational efficiency of health services in Mozambique.

96

00:12:18,800 --> 00:12:31,200

So, in 2006 there were only five clinics offering HIV testing services

97

00:12:31,760 --> 00:12:38,640

and the average travel distance between a neighbourhood and a clinic is about 17.7 kilometres.

98

00:12:39,440 --> 00:12:48,720

And in the following three years there were 27 additional clinics selected to provide HIV

99

00:12:48,720 --> 00:13:02,480

testing services. The average travel distance was 6.3 kilometres. So, we used PMP to select

100

00:13:02,480 --> 00:13:08,720

those additional 27 clinics and we demonstrated that if those additional

101

00:13:08,720 --> 00:13:19,840

27 clinics were selected by PMP the average travel distance could be reduced to 4.7 kilometres.

102

00:13:22,960 --> 00:13:29,520

So, the second model is Location Set Covering Problem. So, it is

103

00:13:30,080 --> 00:13:34,160

to find the minimum number of facilities and their locations

104

00:13:34,160 --> 00:13:38,400

such that all neighbourhoods are covered within the maximal distance or time standard.

105

00:13:39,840 --> 00:13:47,760

So, in the example here we assume the facilities are schools and the demands are neighbourhoods.

106

00:13:49,280 --> 00:13:56,880

So, we also define the impedance cutoff. You know this can be distance or travel time.

107

00:13:56,880 --> 00:14:02,640

So here we find a circular catchment area for each school. So, a

108

00:14:03,760 --> 00:14:12,000

school only can serve the neighbourhoods that are within this catchment area. So, the problem would be

109

00:14:14,400 --> 00:14:20,560

at least how many schools we need to serve all the neighbourhoods here.

110

00:14:21,280 --> 00:14:29,840

So here we can see at least we need four schools to serve all the neighbourhoods.

111

00:14:31,760 --> 00:14:38,960

So, this is the mathematical formulation of the LSCP.

112

00:14:38,960 --> 00:14:46,560

We use the notation as before and here we have two additional notations. a_{ij} indicates whether

113

00:14:49,440 --> 00:14:56,480

demand area i is served by a potential facility j and N_i is the set of

114

00:14:56,480 --> 00:15:03,040

all facilities that can provide proper services to demand area i.

115

00:15:04,240 --> 00:15:10,640

So, the decision variable X_j which indicates whether a facility is located or not.

116

00:15:13,600 --> 00:15:20,720

So, we look at the model on the right-hand side. The objective is to minimise the total number of

117

00:15:20,720 --> 00:15:28,480

sited facilities, and constraint one ensures each demand area is provided suitable coverage.

118

00:15:29,360 --> 00:15:33,760

So here it means for all the facilities

119

00:15:34,560 --> 00:15:45,520

that can serve demand area i, at least one of those facilities must be selected and constraint two requires

120

00:15:45,520 --> 00:15:53,840

the value of X_j only can be zero or one. We either site facility *j* or not.

121

00:16:00,400 --> 00:16:11,680

So, this is an example of optimising fire station locations in Nanjing, China.

122

00:16:13,280 --> 00:16:20,400

The first map on the left-hand side shows the existing 19 fire stations in

123

00:16:20,400 --> 00:16:24,560

the urban area of Nanjing. So here if we use

124

00:16:26,640 --> 00:16:30,800

four kilometres as a service standard. We can see

125

00:16:33,360 --> 00:16:43,360

there are some areas that cannot be covered by existing fire services. So next we consider two

126

00:16:43,360 --> 00:16:50,800

scenarios. First one, we assume there were no existing fire stations and

127

00:16:51,360 --> 00:17:01,280

we look at at least how many fire stations we need to serve the entire area. And we use LSCP

128

00:17:01,280 --> 00:17:09,680

and it shows that we need at least 22 fire stations to serve all the urban area.

129

00:17:11,680 --> 00:17:18,720

And the second scenario, we consider the existing fire stations. Given existing 19 stations,

130

00:17:18,720 --> 00:17:25,760

at least how many additional fire stations we need to serve the whole area, and the LSCP

131

00:17:27,120 --> 00:17:34,080

indicates we need at least an additional 13 fire stations to serve all the area.

132

00:17:42,480 --> 00:17:48,800

So, the last model is the Maximal Covering Location Problem.

133

00:17:49,600 --> 00:17:56,560

It is to locate a pre-specified number of facilities such that coverage within a maximal service

134

00:17:57,440 --> 00:18:07,280

distance or time is maximised. So, in the example here we assume the facilities are fire

135

00:18:07,280 --> 00:18:16,240

stations and demands are the neighbourhoods. So, each neighbourhood can have an associated weight, W.

136

00:18:16,240 --> 00:18:23,680

This weight can represent the population in each neighbourhood. And we also define

137

00:18:23,680 --> 00:18:30,160

the service area for the fire station so we can assume the response time

138

00:18:30,160 --> 00:18:39,360

for fire station is like four minutes or five minutes. And again, here we define a circular

139

00:18:39,360 --> 00:18:46,400

service area for each fire station to represent the service coverage. So, each fire

140

00:18:46,400 --> 00:18:53,120

station only can serve the neighbourhoods within this circular service area.

141

00:18:54,800 --> 00:19:00,400

So again, we specify the number of facilities, the fire stations, to be located.

142

00:19:00,400 --> 00:19:06,560

It's four. So, we're going to site four fire stations. So, the problem here would be

143

00:19:07,600 --> 00:19:14,880

where should we site those four fire stations so that the total covered (or served) population (of the neighbourhoods)

144

00:19:14,880 --> 00:19:21,040

is maximised.

145

00:19:21,040 --> 00:19:27,600

So, this is the Maximal Covering Location Problem.

146

00:19:34,160 --> 00:19:38,800

This is the mathematical formulation of the MCLP.

147

00:19:40,400 --> 00:19:51,120

Again, we use the same notation as before: demand area i and the set of facilities that serve the demand area i, N_i; facility location j; w_j is the amount of demand in area i.

148

00:20:02,320 --> 00:20:13,520

So, p is the number of facilities to be located. And here we have two types of decision variables

149

00:20:13,520 --> 00:20:21,200

X_j and Y_i. So X_j again indicates whether the facility is located or not and Y_i indicates

150

00:20:21,200 --> 00:20:29,280

if area *i* is provided coverage or not. So, look at the model on the right-hand side,

151

00:20:30,000 --> 00:20:38,240

the objective is to maximize the total covered demand. So, the w_i represents the

152

00:20:38,240 --> 00:20:45,920

demand in area *i*. And constraint one accounts for demand area provided suitable coverage.

153

00:20:46,560 --> 00:20:57,040

So, if the value of Y_i is one which means the demand area *i* is provided coverage, for all

154

00:20:57,040 --> 00:21:02,640

the facilities

155

00:21:02,640 --> 00:21:09,120

that can provide service to demand area *i*, at least one of those facilities should be located.

156

00:21:10,880 --> 00:21:18,240

If the value of Y_i is zero, so it doesn't matter whether those facilities are sited or not.

157

00:21:20,800 --> 00:21:26,000

Constraint two indicates the total number of facilities to be located

158

00:21:26,000 --> 00:21:35,040

is *p*. So again, constraint three requires the values of X_j and Y_i only can be zero or one.

159

00:21:40,320 --> 00:21:47,120

So far, we've looked at three types of classical facility location models.

160

00:21:47,920 --> 00:21:54,080

So, we looked at the mathematical formulations. So why GIS is important here? Why we

161

00:21:54,080 --> 00:22:03,600

need GIS? So, GIS is important in at least two aspects. The first is spatial representation. So

162

00:22:03,600 --> 00:22:10,640

we can use spatial data to represent facilities and demands. As we know in GIS, we

163

00:22:10,640 --> 00:22:16,480

can represent spatial data using vectors or rasters. We can represent spatial

164

00:22:16,480 --> 00:22:22,960

objects as points, lines and polygons and we can also use grids,

165

00:22:24,000 --> 00:22:29,280

regular grids to represent spatial objects. So here, for example, for a particular region,

166

00:22:29,280 --> 00:22:35,360

we can represent it as a boundary, a set of points or a set of grids.

167

00:22:37,200 --> 00:22:45,600

Second, GIS also can help assess service coverage. So, in practice you will

168

00:22:45,600 --> 00:22:51,440

usually use the travel distance or travel time to represent service coverage. And,

169

00:22:51,440 --> 00:23:01,520

as we know distance or travel time is very easy to calculate within a GIS environment. And also, the

170

00:23:01,520 --> 00:23:07,840

definitions of service coverage can be different depending on the spatial representation of demands.

171

00:23:08,560 --> 00:23:16,480

So, for example, here we have a demand area which can be represented by three points.

172

00:23:16,480 --> 00:23:25,760

The red points here, all represented by a set of three polygons and three grids. So, if this demand,

173

00:23:28,400 --> 00:23:30,880

if this model is represented by three points,

174

00:23:31,440 --> 00:23:38,800

we can see we only need two facilities to cover all those three points, to cover this demand.

175

00:23:40,000 --> 00:23:47,840

And if this demand is represented by three grids, by polygons. And

176

00:23:49,360 --> 00:23:56,480

we can see we need at least three facilities to cover this area, this demand.

177

00:23:57,440 --> 00:24:03,520

So, the solutions of facility location problems can be different,

178

00:24:03,520 --> 00:24:09,760

which depends on the spatial condition of demand. It also can happen to the facilities.

179

00:24:09,760 --> 00:24:17,840

It also can represent the facility as a point or a polygon. So, this is the role of GIS.

180

00:24:22,000 --> 00:24:28,320

So next we will look at the location-allocation analysis function in ArcGIS.

181

00:24:29,600 --> 00:24:34,960

So, although we mentioned we can use different spatial data to represent facilities and the

182

00:24:34,960 --> 00:24:42,640

demand. But in ArcGIS all facilities and demands are represented by points. You will

183

00:24:42,640 --> 00:24:49,600

look at that in the tutorial. So here facilities are supposed to provide goods and services, and

184

00:24:49,600 --> 00:24:57,440

demand points to consume goods and services. And the location-allocation analysis function in

185

00:24:57,440 --> 00:25:05,360

ArcGIS is to help you to choose a subset of facilities from a set of candidate facilities

186

00:25:06,000 --> 00:25:15,600

and this selection will be subject to underlying constraints which is defined as the interaction

187

00:25:15,600 --> 00:25:22,800

between facilities and demand points. So, this function in ArcGIS will

188

00:25:23,440 --> 00:25:28,880

simultaneously locate facilities and

allocate demand points to the facilities.

189

00:25:31,680 --> 00:25:38,320

In total you can solve seven different types of problems in ArcGIS in this module (in the location-allocation module).

190

00:25:38,320 --> 00:25:45,680

We will look at those different problems one by one.

191

00:25:47,120 --> 00:25:52,800

So, minimise impedance - this is actually the previous problem we introduced before.

192

00:25:54,960 --> 00:25:59,760

Facilities are located such that the sum of all weighted costs between demand points and

193

00:25:59,760 --> 00:26:06,800

solution facilities is minimised. You can see here the objective is to minimise the total

194

00:26:06,800 --> 00:26:15,600

weighted costs. Traditionally this model is used to locate warehouses because it can reduce

195

00:26:15,600 --> 00:26:21,760

the overall transportation costs of delivering goods to outlets. So, the transportation costs are

196

00:26:21,760 --> 00:26:28,720

usually associated with the travel time or travel distance and also related to the

197

00:26:30,880 --> 00:26:32,480

the amount of demand needed at the outlet.

198

00:26:35,200 --> 00:26:42,560

It is also commonly used for locating public sector facilities,

199

00:26:42,560 --> 00:26:48,320

such as libraries, regional airports and museums and so on.

200

00:26:50,000 --> 00:26:55,840

So, by minimising the total weighted transportation costs,

201

00:26:56,480 --> 00:27:03,040

as I said before, we assume that we can improve the system efficiency and

202

00:27:03,040 --> 00:27:11,120

actually we are trying to maximise the accessibility to achieve social equity.

203

00:27:13,280 --> 00:27:16,720

So, in the example on the right hand side we can see we

204

00:27:17,280 --> 00:27:24,560

locate three facilities to serve five demand points.

205

00:27:29,600 --> 00:27:43,120

So, the second model is the Maximise Coverage. This is similar to the MCLP

206

00:27:43,120 --> 00:27:49,840

(Maximal Coverage Location Problem) as we introduced before. So, we can see here

207

00:27:51,040 --> 00:27:55,920

the facilities are located such that as many demand points as possible are allocated to

208

00:27:55,920 --> 00:28:03,920

solution facilities within the impedance cutoff. For this problem, we have an impedance cutoff

209

00:28:03,920 --> 00:28:11,120

which defines the surface area for each facility and the objective is to cover or to serve as

210

00:28:11,120 --> 00:28:17,600

many demand points as possible. So, this type of problem is often used to locate fire stations,

211

00:28:17,600 --> 00:28:24,800

police stations and emergency centres. As we know the emergency services always have

212

00:28:24,800 --> 00:28:30,560

(usually have) a specified response time like, you know, for fire stations (for fire services), we require

213

00:28:30,560 --> 00:28:37,680

a response time (it's) like 4 minutes or 5 minutes. So, for example on the right-hand side we can see

214

00:28:39,360 --> 00:28:48,240

here we site three facilities and those three facilities serve

215

00:28:49,680 --> 00:28:56,320

five demand points, and one demand point on the top is outside of the service area

216

00:28:56,320 --> 00:29:01,280

of all those three facilities, so it is not allocated (so it is not served).

217

00:29:07,440 --> 00:29:08,880

The third problem is

218

00:29:11,120 --> 00:29:18,000

Maximise Capacitated Coverage. So, it is similar to the previous problem Maximise Coverage.

219

00:29:18,640 --> 00:29:24,960

In addition, it has an additional requirement (constraint) here.

220

00:29:26,560 --> 00:29:33,200

So, each facility will have a capacity, so the weighted demand allocated to a facility cannot exceed

221

00:29:33,200 --> 00:29:42,240

the facility's capacity. So again, the objective is to cover as many demand points as possible

222

00:29:43,280 --> 00:29:51,120

and also, each facility has a service area defined (pre-specified service area). We

223

00:29:51,120 --> 00:29:56,880

have our impedance cutoff here which can be represented by travel time or travel distance.

224

00:29:58,640 --> 00:30:06,160

So, this type of problem can be used for creating territories that encompass a given number of

225

00:30:06,160 --> 00:30:14,560

people or businesses, or to locate hospitals or other medical facilities with a limited number of

226

00:30:14,560 --> 00:30:20,640

beds or patients who can be treated, or to locate warehouses whose inventory isn't assumed to be

227

00:30:20,640 --> 00:30:30,880

unlimited. So again, for the example on the right-hand side we site three facilities and each facility

228

00:30:30,880 --> 00:30:38,800

only serves one demand point. So, although there is a demand point on the bottom here, this

229

00:30:38,800 --> 00:30:45,520

demand point is within the service area of a yellow facility, but

230

00:30:46,480 --> 00:30:53,440

it is not allocated to that facility because that facility does not have enough capacity

231

00:30:53,440 --> 00:31:03,040

to serve two demand points. Although that demand point is within

232

00:31:03,040 --> 00:31:10,000

the service area, but it is not served. So, in this case we consider the facility's capacity.

233

00:31:13,760 --> 00:31:22,240

The next: Minimise Facilities. So, this is equivalent to the LSCP (the Location Set Cover

234

00:31:22,240 --> 00:31:28,720

Problem) we introduced before. So, we can see here the objective is to minimise the number of

235

00:31:28,720 --> 00:31:34,400

facilities. So, the number of facilities required to cover demand points is minimised.

236

00:31:35,120 --> 00:31:41,760

But different from the LSCP, so in this problem, in this model, we also require

237

00:31:41,760 --> 00:31:48,720

(we also have) impedance cutoff here. So, facilities are located as such that as many demand points

238

00:31:48,960 --> 00:31:53,680

as possible are allocated to solution facilities within the impedance cutoff.

239

00:31:56,160 --> 00:32:06,720

And this type of model is often used when the cost of building

240

00:32:06,720 --> 00:32:12,000

facilities is not a limiting factor. So, we don't need to consider the cost

241

00:32:12,000 --> 00:32:19,280

here. And it is often used to choose school bus stops when students are required to walk a

242

00:32:19,280 --> 00:32:25,680

certain distance before another school bus stop is provided closer to the student's residence.

243

00:32:27,680 --> 00:32:45,520

And so, for the example here, so we can see there is a demand point on the top right. It is outside of the service area of those

244

00:32:45,520 --> 00:32:54,240

two yellow facilities, so it cannot be covered, and only those four demand points

245

00:32:54,240 --> 00:33:01,520

are within the service areas of those two yellow facilities. So, we want to,

246

00:33:04,080 --> 00:33:10,560

so we select those two facilities to serve all those four demand points. So, we need

247

00:33:10,560 --> 00:33:18,800

at least two facilities to serve all the possible demand points that can be covered.

248

00:33:23,200 --> 00:33:25,840

So next Maximise Attendance.

249

00:33:28,000 --> 00:33:33,520

The facilities are chosen such that as much demand weight as possible is allocated to facilities

250

00:33:33,520 --> 00:33:38,160

while assuming the demand weight decreases in relation to the distance between the facility

251

00:33:38,160 --> 00:33:43,920

and the demand point. Again, we want to cover as much demand weight as possible.

252

00:33:45,120 --> 00:33:48,720

So, each one has a weight here and actually the demand weight

253

00:33:49,280 --> 00:33:55,840

will decrease in relation to the distance between the facility and demand point. So, it assumes that

254

00:33:55,840 --> 00:34:01,520

the further people have to travel to reach your facility, the less likely they are to use it.

255

00:34:03,280 --> 00:34:09,360

So public transit bus stops are often chosen with

256

00:34:09,360 --> 00:34:18,240

the help of this type of model. This model also can be beneficial

257

00:34:18,240 --> 00:34:25,280

to specialty stores that have little or no competition, general retailers, and restaurants

258

00:34:25,280 --> 00:34:31,360

that don't have the data on competitors and some businesses including coffee shops, fitness centres

259

00:34:33,520 --> 00:34:36,560

and electronic stores and so on.

260

00:34:38,000 --> 00:34:47,840

So, for the example here on the right-hand side you can see we only site one facility and it serves

261

00:34:47,840 --> 00:34:53,840

five demand points and there are two demand points outside of the service area.

262

00:34:54,640 --> 00:35:03,520

and the pie chart here indicates how much of total demand is captured by the facility.

263

00:35:05,520 --> 00:35:13,120

So, we can see here, if a demand point is closer to the facility the percent of

264

00:35:13,680 --> 00:35:23,360

weight captured by the facility will be higher. So, if the demand is further away from the

265

00:35:24,480 --> 00:35:32,080

facility the percent of demand captured by the facility will be lower. So that

266

00:35:33,760 --> 00:35:43,840

is what we mean here: demand weight decreases in relation to the distance between the facility and the demand point.

267

00:35:44,800 --> 00:35:51,520

Maximise Market Share. So here again we specify the number of facilities we are going to site

268

00:35:52,640 --> 00:36:00,640

(and) in order to maximise the allocated demand in the presence of competitors. So, in this model

269

00:36:00,640 --> 00:36:06,960

we have two types of facilities. One is the facility we are going to site

270

00:36:06,960 --> 00:36:14,800

and the other is the facilities of the competitors. The goal is to capture as much

271

00:36:14,800 --> 00:36:22,880

of the total market share as possible with a given number of facilities. And this type of problem

272

00:36:22,880 --> 00:36:30,240

actually uses the Huff model which is also known as a gravity model or spatial interaction model.

273

00:36:31,200 --> 00:36:38,560

So, we don't explain the Huff model here, but you can find the explanation in any

274

00:36:38,560 --> 00:36:46,800

Basic GIS textbook or spatial analysis book. So large discount stores typically use

275

00:36:46,800 --> 00:36:55,840

Maximise Market Share to locate a finite set of new stores. So, for example, here on the

276

00:36:55,840 --> 00:37:00,320

right hand side, so we can see there is a demand point near the middle of the graph.

277

00:37:01,520 --> 00:37:07,440

So, this pie chart indicates the percent of demand weight captured by the facility of same

278

00:37:07,440 --> 00:37:14,960

colour. And we can see this demand point is within the service area of both yellow facility and blue

279

00:37:14,960 --> 00:37:22,000

facility. And because it is closer to the yellow facility, so we can see the percent of demand weight

280

00:37:22,000 --> 00:37:28,000

captured by the yellow facility is higher than that captured by the blue facility.

281

00:37:30,960 --> 00:37:39,920

So, Target Market Share, it chooses the minimum number of facilities necessary to capture a

282

00:37:39,920 --> 00:37:45,920

specific percentage of the total market share in the presence of competitors. So, in this problem again

283

00:37:45,920 --> 00:37:53,360

we have competitors and we need to specify the percentage of the total market share that we want to cover.

284

00:37:54,880 --> 00:38:01,840

So again, in this case budgets are not a concern. We need to

285

00:38:02,960 --> 00:38:10,720

identify the minimum number of facilities we need to achieve this desired

286

00:38:10,720 --> 00:38:18,080

market share. So large discount stores typically use the Target Market Share problem type

287

00:38:18,080 --> 00:38:23,680

when they want to know how much expansion would be required to reach a certain level of the market share,

288

00:38:24,800 --> 00:38:29,360

and also what strategy would be needed just to maintain their current market share

289

00:38:30,000 --> 00:38:38,400

given the introduction of new competing facilities. So again, here we can see we have two types of

290

00:38:38,400 --> 00:38:49,040

facilities. One is represented by yellow; one is in blue. And there are two demand points. They

291

00:38:49,040 --> 00:38:59,840

are within the service area of both yellow and the blue facilities. And we can see for the demand point on the right which can be served by both blue facility and yellow facility.

292

00:39:09,920 --> 00:39:16,560

So, because it has the, so the distance of this demand point to the blue,

293

00:39:17,120 --> 00:39:21,280

to the nearest blue facility and the yellow facility is the same. So, we can see

294

00:39:21,280 --> 00:39:30,160

the demand weight of this point is evenly shared by the blue facility and the yellow facility.

295

00:39:31,440 --> 00:39:37,840

So, demand weight is evenly split between the yellow facility and the blue facility.

296

00:39:40,400 --> 00:39:48,880

So that's the seven different types of location-allocation

297

00:39:48,880 --> 00:39:57,280

analysis provided by ArcGIS. So next I will briefly overview the tutorial we provided

298

00:39:58,400 --> 00:40:04,960

for this webinar. So, you can find the data and the tutorial notes from the UBDC website.

299

00:40:07,760 --> 00:40:14,080

So, in this tutorial we are going to look at some recycling facilities for

300

00:40:14,080 --> 00:40:18,240

example glass recycling bins in Glasgow city centre. And

301

00:40:19,600 --> 00:40:27,440

we need two types of data. The first is road network. We use data from the OpenStreetMap.

302

00:40:27,440 --> 00:40:34,240

The second type of data is the centroids of Output Areas of Glasgow. So, the Output

303

00:40:34,240 --> 00:40:42,960

Areas is Census Geography in Scotland. It is a type of polygon (defined),

304

00:40:44,560 --> 00:40:51,920

it is the type of polygon used in the Census. So, we use the centroids of the Output Areas

305

00:40:51,920 --> 00:40:58,320

to represent both facilities and demands. So here we use the same set of points to represent the

306

00:40:58,320 --> 00:41:08,120

facilities and demands. And each demand point has a weight. So, this weight represents the number of people living in

307

00:41:08,320 --> 00:41:15,360

each area. So, we will use ArcGIS location-allocation analysis (the location-allocation analysis function in ArcGIS)

308

00:41:16,000 --> 00:41:26,400

to look at three examples of facility location problems here and

309

00:41:26,400 --> 00:41:34,720

on this map all those green dots actually represent, those green dots are centroids

310

00:41:34,720 --> 00:41:41,600

of Output Areas in Glasgow city centre. We will use them to represent the

311

00:41:41,600 --> 00:41:48,800

potential sites for recycling facilities and all the demand areas.

312

00:41:50,960 --> 00:41:54,560

So, for the first example we will look at the minimise facilities problem.

313

00:41:55,680 --> 00:42:03,520

And we assume everyone should be within a 500-metre distance over a recycling bin. So here the impedance cutoff

314

00:42:03,520 --> 00:42:10,800

is 500 metres. And we want to know what is the smallest number of bins we could use to do this

315

00:42:10,800 --> 00:42:21,840

and where should we place them to find the minimum number of bins to serve all the neighbourhoods,

316

00:42:21,840 --> 00:42:31,680

with 500 metre impedance cutoff. So, the map on the right-hand side

317

00:42:31,680 --> 00:42:41,200

shows the result. So, the red squares indicate the locations

318

00:42:41,200 --> 00:42:48,160

we selected. We will site recycling bins at those locations and we also can see,

319

00:42:49,040 --> 00:43:00,240

the neighbourhood centroids (the demand points) are allocated

320

00:43:00,240 --> 00:43:07,680

to its nearest facility. So that's why we call this problem location-allocation. We locate

321

00:43:07,680 --> 00:43:13,920

recycling bins and then we allocate demand points to the nearest facilities.

322

00:43:16,320 --> 00:43:25,840

So, the second example: maximise coverage. So, we assume the service distance is

323

00:43:25,840 --> 00:43:35,440

500 metres but in this case we will have a budget to have four recycling facilities. So, we

324

00:43:35,440 --> 00:43:40,720

only can site four facilities. We want to know where to put them so we can cover the

325

00:43:41,520 --> 00:43:49,520

maximum demand (in this case we can cover the maximum population). So where should we place

326

00:43:49,520 --> 00:43:58,480

them? The map on the right-hand side shows the result. You can see we sited four

327

00:43:59,360 --> 00:44:06,000

facilities here and only the demand points within their service area can be served.

328

00:44:06,800 --> 00:44:12,560

There are many demand points here that cannot be served by those four facilities.

329

00:44:15,600 --> 00:44:26,320

So, the third example is to minimise impedance. So, this is equivalent to the p-median

330

00:44:28,240 --> 00:44:37,920

problem to minimise the weighted distance from each demand node to each recycling

331

00:44:37,920 --> 00:44:44,720

facility. So here p is four, so we want to site four facilities. So, the total weighted

332

00:44:44,720 --> 00:44:52,960

travel distance or travel time is minimised. So, in this case we don't have impedance cutoff,

333

00:44:56,400 --> 00:45:01,200

and the map on the right-hand side shows the result where we site four recycling bins, and each demand point is allocated to its nearest facility.

334

00:45:11,680 --> 00:45:18,691

So that is all and thanks for your attention. So, I think next I will take questions.

335

00:45:30,400 --> 00:45:34,080

[Weicong Luo] Yes, I have received some

336

00:45:34,080 --> 00:45:40,880

questions and I will represent the participants to ask Jing. And the

337

00:45:41,600 --> 00:45:48,640

the first question is what is the potential for GIS in retail analysis and site location research?

338

00:45:51,520 --> 00:46:03,120

[Jing Yao] So actually I think I covered some of this in my presentation. So, as you can see

339

00:46:03,920 --> 00:46:08,640

the title of this webinar is GIS-based facility location analysis for the public and

340

00:46:08,640 --> 00:46:18,160

private sectors. Although I present the different mathematical models you know, but

341

00:46:19,200 --> 00:46:25,120

essentially facility location problem is a spatial problem. As I said,

342

00:46:26,320 --> 00:46:36,400

basically, we can use GIS to represent the facilities and demands. So, we need

343

00:46:36,400 --> 00:46:43,600

to use certain spatial data to represent facilities and demands and this can be done in GIS,

344

00:46:44,880 --> 00:46:51,360

using points, lines, polygons, and grids to represent facilities and demand.

345

00:46:51,920 --> 00:46:59,360

And also, we can use GIS/spatial analysis to assess spatial relationships,

346

00:46:59,360 --> 00:47:08,880

for example, to measure distance, to measure travel time. And this is the role of GIS. And also, we can see the

347

00:47:08,880 --> 00:47:13,840

definitions of service demand can be different depending on the spatial representation.

348

00:47:15,040 --> 00:47:19,840

So, this means that you might have different, so for example for a specific location problem,

349

00:47:19,840 --> 00:47:25,840

you might have different solutions if you use different spatial representations.

350

00:47:28,320 --> 00:47:38,000

So, this is just two basic functions, the basic roles of GIS in facility location modelling.

351

00:47:38,000 --> 00:47:44,960

GIS also can be used to analyse some

352

00:47:44,960 --> 00:47:48,240

other spatial relationships. Like you know whether

353

00:47:48,240 --> 00:47:54,560

the two facilities like whether two demand polygons are adjacent to each other,

354

00:47:54,560 --> 00:48:00,800

whether the service/market share/the market areas are overlapped. So, we can see

355

00:48:03,920 --> 00:48:07,360

here we can see the market area for the two facilities

356

00:48:08,720 --> 00:48:15,520

are overlapped like for the yellow facility, the blue facility and in GIS it's also

357

00:48:15,520 --> 00:48:19,440

easy to evaluate, you know, what is the area of the overlapped area.

358

00:48:30,880 --> 00:48:37,200

I think that's the main role of GIS in facility location analysis.

359

00:48:39,600 --> 00:48:47,840

[Weicong Luo] And the next question is what is your view about current GIS vendor solutions to

360

00:48:48,400 --> 00:48:56,240

retail site evaluations and sales projections are really crude and will always build a real

361

00:48:56,800 --> 00:49:01,840

unreliable result in comparisons with customer GIS and modelings?

362

00:49:05,600 --> 00:49:14,960

[Jing Yao] I think it has some point here. So, we can see in ArcGIS we have these

363

00:49:14,960 --> 00:49:22,320

seven different location-allocation problem types. So, it means you only can use ArcGIS

364

00:49:22,320 --> 00:49:32,480

to solve these seven different types of problem and if you use ArcGIS you must follow the requirements set by ArcGIS. For example, in the tutorial we mentioned we need a road network here.

365

00:49:41,120 --> 00:49:45,600

So, if you do the location analysis in ArcGIS,

366

00:49:45,600 --> 00:49:56,160

you need a road network. Although sometimes you might

367

00:49:57,600 --> 00:50:06,800

not need that. But in ArcGIS, you must provide a road network in the ArcGIS data format.

368

00:50:10,400 --> 00:50:20,240

And also, all the facilities and demand only can be represented as points. This is a limitation. And also, it has a limited number of

369

00:50:20,880 --> 00:50:27,280

constraints. For example, we can see here for the MCLP or the

370

00:50:28,640 --> 00:50:35,920

LSCP or the p-median problem here. So, all the seven problems you find in ArcGIS it has a limited number

371

00:50:35,920 --> 00:50:42,800

of constraints and in practice you might want to add other constraints. So, it is

372

00:50:42,800 --> 00:50:51,040

another limitation. But you only can put a limited number of constraints in ArcGIS to solve a

373

00:50:51,040 --> 00:50:58,320

particular problem. So, you cannot add your additional constraints

374

00:50:58,320 --> 00:51:10,400

here, so this is a limitation. So, you only can use the ArcGIS location-allocation analysis module

375

00:51:11,920 --> 00:51:20,960

To, maybe to, get an overview of your problem or maybe it can provide some

376

00:51:22,080 --> 00:51:29,280

candidate solutions to a problem. If you want to solve your problem in practice I

377

00:51:29,280 --> 00:51:38,400

think usually and very often we need to customise the modeling. So, you need a more

378

00:51:38,400 --> 00:51:46,320

complicated model, for example, like I listed here. So, in the classical MCLP you only have three

379

00:51:46,320 --> 00:51:52,400

constraints and for your own problem you might want to add more constraints. Maybe

380

00:51:52,400 --> 00:51:58,720

you will have 10 or more constraints. Then you need to write your own model using some

381

00:51:58,720 --> 00:52:09,840

scripts and then solve the model using some optimization software, then you can represent your result

382

00:52:09,840 --> 00:52:16,560

in GIS software to make a map. So that is actually what we typically

383

00:52:16,560 --> 00:52:28,080

do. Again, I mean, but ArcGIS is still useful, you know, to give you an idea that what location-allocation analysis is,

384

00:52:31,680 --> 00:52:35,200

and what kind of models we can use and for what kinds of purposes.

385

00:52:35,760 --> 00:52:42,000

And you can use this model to solve simple problems and for more

386

00:52:43,840 --> 00:52:50,480

complex problems I think definitely you need to write your own customised the model.

387

00:52:51,200 --> 00:52:58,320

And that's what we do for the advanced facility location modeling. So, we

388

00:52:58,320 --> 00:53:06,720

usually write our own, build upon those classical models, then we write our own models

389

00:53:06,720 --> 00:53:12,640

you know by adding other decision variables and other constraints. Then we can use Python or R to

390

00:53:12,640 --> 00:53:20,240

write our model and solve it by open-source software or by a commercial optimization software.

391

00:53:20,240 --> 00:53:27,760

Then we can present the results in ArcGIS. Of course we also need GIS to prepare the

392

00:53:27,760 --> 00:53:38,160

input data, to prepare the demand data and facility data. So, I agree, yes to some extent

393

00:53:41,600 --> 00:53:48,560

the solutions like provided by the location-allocation analysis in ArcGIS are crude,

394

00:53:53,360 --> 00:53:57,200

Usually, we don't use the results from ArcGIS directly,

395

00:53:57,200 --> 00:54:01,840

we just use that to help us with the decision making.

396

00:54:05,600 --> 00:54:12,320

[Weicong Luo] In my own research, because ArcGIS

397

00:54:12,320 --> 00:54:18,560

just provides some basic models so if you want to research more details or

398

00:54:18,560 --> 00:54:23,360

more accurate models which would transform your spatial data to

399

00:54:23,360 --> 00:54:31,440

the numbers and use the software or Python, R, too, to solve it and return to the spatial data in ArcGIS.

400

00:54:32,960 --> 00:54:39,360

And the last question is, the number of demand and the number of facilities are having to

401

00:54:39,360 --> 00:54:47,840

be the same? [Jing Yao] No, no, definitely not. So, in our tutorial,

402

00:54:48,960 --> 00:54:54,720

so we just use this as an example, so it's easier, we only need to prepare one set of data so to

403

00:54:54,720 --> 00:55:00,080

represent both facilities and demand points. But in reality, you can have two different sets of data,

404

00:55:00,080 --> 00:55:07,120

one is facilities, one is for demand points.

405

00:55:07,680 --> 00:55:13,520

You don't have to have the same number of facilities and demand points. They can be different.

406

00:55:17,680 --> 00:55:19,840

[Weicong Luo] More questions?

407

00:55:24,240 --> 00:55:31,280

So, this is all of the questions and we will put our question and answers in writing

408

00:55:31,920 --> 00:55:38,320

and send this to the website because some people are not hearing very clear.

409

00:55:56,480 --> 00:56:03,200

We will also upload the slides to the UBDC website so you can

410

00:56:03,760 --> 00:56:11,120

check the information later and also read the references provided

411

00:56:11,120 --> 00:56:19,280

here. So, for each model, so we have some references, the origin of this model.

412

00:56:22,560 --> 00:56:28,240

And for the empirical studies I also put a link to an article here so you can read the

413

00:56:28,240 --> 00:56:36,720

article to get more information on how we applied the specific location models

414

00:56:38,240 --> 00:56:45,040

to solve the practical problems. Of course, we didn't apply the classical model directly. We

415

00:56:45,040 --> 00:56:52,800

usually would like to, for example, add more constraints here to make it suitable for our

416

00:56:52,800 --> 00:56:59,360

Practical problem. For this one we also can find the article here.

417

00:57:11,440 --> 00:57:16,960

Here you can get other information about data from the UBDC website.

418

00:57:23,280 --> 00:57:32,720

If no more questions I guess we can end then. Thanks for your participation and

419

00:57:34,960 --> 00:57:39,840

we can end today's webinar.